



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>G01B 11/04</b>	<b>A1</b>	(11) International Publication Number: <b>WO 97/35161</b> (43) International Publication Date: 25 September 1997 (25.09.97)
--	-----------	--

(21) International Application Number: PCT/US97/02115

(22) International Filing Date: 12 February 1997 (12.02.97)

## (30) Priority Data:

60/011,524	12 February 1996 (12.02.96)	US
Not furnished	11 February 1997 (11.02.97)	US

(71) Applicant: SARNOFF CORPORATION [US/US]; 201 Washington Road, CN 5300, Princeton, NJ 08543-5300 (US).

(72) Inventors: IRANI, Michal; 11-B Landing Lane, Princeton Junction, NJ 08550 (US). ANANDAN, Pahmanabhan; B-2 Carver Place, Lawrenceville, NJ 08648 (US).

(74) Agents: DiBARTOLOMEO, Anthony, L. et al.; RATNER &amp; PRESTIA, Suite 301, One Westlakes, Berwyn, P.O. Box 980, Valley Forge, PA 19482-0980 (US).

(81) Designated States: CA, JP, KR, MX, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

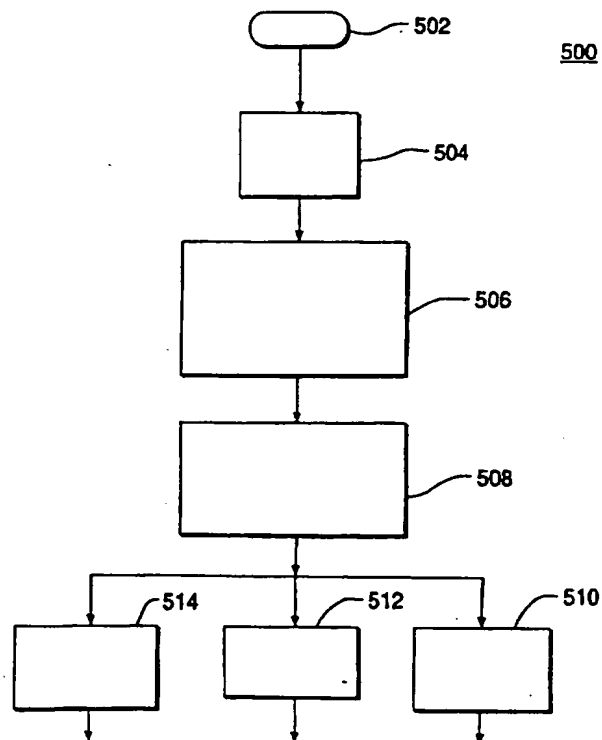
## Published

*With international search report.*  
*With amended claims.*

(54) Title: METHOD AND APPARATUS FOR THREE-DIMENSIONAL SCENE PROCESSING USING PARALLAX GEOMETRY OF PAIRS OF POINTS

## (57) Abstract

A method and apparatus for processing a geometric relationship between the image motion of pairs of points over multiple image frames representing a three-dimensional scene. This relationship is based on the parallax motion of points with respect to an arbitrary planar surface, and does not involve epipolar geometry. A constraint is derived over two frames (504) for any pair of points (506), relating their projective structure (with respect to the plane) based only on their image coordinates and their parallax displacements. Similarly, a 3D-rigidity constraint between pairs of points over multiple frames is derived. Applications disclosed for these parallax-based constraints include recovery of 3D scene structure (512), detection of moving objects in the presence of camera induced motion (510), and synthesis of new camera views based on a given set of views (514). Moreover, the approach can handle difficult situations for 3D scene analysis, e.g., where there is only a small set of parallax vectors, and in the presence of independently moving objects.



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

5

METHOD AND APPARATUS FOR THREE-DIMENSIONAL SCENE  
PROCESSING USING PARALLAX GEOMETRY OF PAIRS OF POINTS

10

FIELD OF THE INVENTION

15 The invention generally relates to image processing systems and, more particularly, to a method and apparatus for processing the parallax geometry of pairs of points within three-dimensional scene.

### BACKGROUND OF THE INVENTION

The analysis of three dimensional scenes from image sequences has a number of goals. These goals include, but are not limited to: (i) the recovery of  
5 3D scene structure, (ii) the detection of moving objects in the presence of camera induced motion, and (iii) the synthesis of new camera views based on a given set of views.

The traditional approach to these types of problems has been to first recover the epipolar geometry between pairs of frames and then apply that  
10 information to achieve the above-mentioned goals. However, this approach is plagued with the difficulties associated with the recovery of the epipolar geometry.

Recent approaches to 3D scene analysis have attempted to overcome some of the difficulties in recovering the epipolar geometry by decomposing the motion  
15 into a combination of a planar homography and residual parallax. The residual parallax motion depends on the projective structure, and the translation between the camera origins. While these methods remove some of the ambiguities in estimating camera rotation, they still require the explicit estimation of the epipole itself, which can be difficult under many  
20 circumstances. In particular, epipole estimation is ill-conditioned when the epipole lies significantly away from the center of the image and the parallax motion vectors are nearly parallel to each other. Also, when there are only a small number of parallax vectors and the scene contains moving objects, these objects incorrectly influence the estimation of the epipole.

25 In general, the treatment of multipoint geometry assumes that the scene is static and relies on the fact that almost all points selected for the shape estimation are known to belong to a single rigid body. In its current form, this class of methods has drawbacks, for example, these methods do not address the problem of shape recovery in dynamic scenes, in particular when the amount of  
30 image motion due to independent moving object is not negligible.

### SUMMARY OF THE INVENTION

The present invention involves a technique for image processing which receives a plurality of two dimensional images representative of a scene,

computes a parallax-related constraint for a pair of points within the plurality of images, applies the parallax-related constraint to a plurality of points within the plurality of images in order to generate information representative of whether a given point within the plurality of images is consistent with the parallax-related  
5 constraint; and uses the generated information for an image processing task related to the received plurality of images.

### BRIEF DESCRIPTION OF THE DRAWINGS

The teachings of the present invention can be readily understood by  
10 considering the following detailed description in conjunction with the accompanying drawings in which:

FIG. 1 depicts a block diagram of a three-dimensional scene analysis system;

FIG. 2 provides a geometric illustration of the planar homography plus  
15 parallax decomposition;

FIG. 3 depicts a geometric illustration of a pairwise parallax constraint;

FIG. 4 depicts a geometric illustration of a scene where epipole estimation is unreliable, but the relative structure constraint can be reliably used to recover relative structure within the scene;

FIG. 5 depicts a flow diagram of a routine which utilizes the parallax  
20 constraint;

FIGS. 6 and 7 provide an illustration of parallax geometry and the dual of the epipole; and

FIGS. 8a-g depict a series of images used and produced during shape  
25 recovery that relies on a single parallax vector.

FIGS. 9a-b illustrates reliable detection of 3D motion inconsistency with sparse parallax information using a ball and tree.

FIGS. 10a-f are a series of images illustrating moving object detection relying on a single parallax vector.

FIGS. 11a-f are a series of images, like FIGS. 10a-f, illustrating moving  
30 object detection relying on a single parallax vector.

To facilitate understanding, identical reference numerals have been used, where possible, to designate identical elements that are common to the figures.

## DETAILED DESCRIPTION

### Overview

The present invention uses, in various image processing tasks, the geometric relationships between the parallax displacements of two or more points within two or more images. The invention applies these relationships to various image processing tasks such as (i) the recovery of 3D scene structure, (ii) the detection of moving objects in the presence of camera induced motion, and (iii) the synthesis of new camera views based on a given set of views.

An important advantage of the present invention is the ability to effectively operate in difficult image processing situations (e.g., when there are a small number of parallax vectors, when the epipole estimation is ill-conditioned, and in the presence of moving objects). The present invention does not require the recovery of epipoles during processing; therefore, it applies to situations when the accurate recovery of the epipole is difficult. Moreover, the present techniques for 3D scene analysis are applicable when only a small set of parallax vectors is available. In fact, the planar parallax of a single point can be used as a reference to recover the structure of the entire scene, and to determine whether other points belong to the static scene or to a moving object.

The results presented herein are expressed in terms of the residual parallax displacements of points after canceling a planar homography. The decomposition of image motion into a homography plus parallax was shown to be more robust, yet more general, than the decomposition into rotation plus translation. Techniques for estimating the planar homography from pairs of images is described in J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation" *European Conference on Computer Vision*, pages 237-252, Santa Margarita Ligure, May 1992, which is herein incorporated by reference.

In the present invention, a parallax-based structure constraint is derived that relates the projective structure of two points to their image positions and their parallax displacements. By eliminating the relative projective structure of a pair of points between three frames, a constraint, referred to as the parallax-based rigidity constraint, on the parallax displacements of two points moving as a rigid object over those frames is derived.

Also presented below is an alternative way of deriving the parallax-based rigidity constraint. In the alternative derivation, the constraint is determined geometrically rather than algebraically. Doing so leads to a simple and intuitive geometric interpretation of the multiframe rigidity constraint and to the  
5 derivation of a dual point to the epipole.

Examples of applications of these parallax-based constraints to solving three important problems in 3D scene analysis are also described. The applications include: (i) the recovery of 3D scene structure, (ii) the detection of moving objects in the presence of camera induced motion, and (iii) the synthesis  
10 of new camera views based on a given set of views.

Finally, the generalization of the constraint to full image motion by including the planar homography component is described.

#### Exemplary Embodiment

15 Turning to the figures, FIG. 1 depicts a block diagram of a three-dimensional scene analysis system 100 suitable for implementing of the present invention. The system contains an image source 102, a computer system 104, one or more output devices 106 and one or more input devices 108. The image source 102 can be a video camera, an infrared camera, or some other  
20 sensor that generates a series of two-dimensional images representing a scene. Alternatively, the image source can be a storage device such as a video tape recorder, disk drive or some other means for storing sequential images representing a scene. The system generally processes digital images; therefore, if the image source produces analog images, a digitizer (not shown) is used  
25 between the image source and the computer system.

The general purpose computer 104 facilitates image processing, scene analysis and image display. Specifically, the computer system contains a data buffer 110, a central processing unit (CPU) 112, support circuitry 114, random access memory (RAM) 116, read only memory (ROM) 118, and a display  
30 driver 120. Additionally, a user interacts with the computer system through one or more input devices 108 such as a keyboard, mouse, trackball, touchpad, and the like. Also, the computer systems displays the images and various graphical interface displays (screens) on the output display device 106 such as a computer monitor. Alternatively, the computer system may also interact with other

output display devices such as a printer to provide a "hard copy" of any display that appears on the computer monitor.

The data buffer 110 provides data rate equalization (frame buffering) between the image source and the CPU. Typically, this buffer is a first-in, first-out (FIFO) buffer. Such buffers are typically used to provide a constant data rate to the CPU while providing flexibility in the data rates that can be generated by an image source.

The CPU 112 is typically a general purpose processor such as a PowerPC, Pentium, or some other generally available processor. PowerPC is a registered trademark of International Business Machines of Armonk, New York and Pentium is a registered trademark of Intel Corporation of Santa Clara, California. Since the software implementation of the present invention is not required to execute on any specific processor, the routines of the present invention can be executed upon any type of processor or combination of processors in a parallel processing computer environment. In addition, rather than using a general purpose computer, the scene analysis may be accomplished within a real-time image processor.

The CPU 112 operates in conjunction with various other circuits such as RAM 116, ROM 118 and support circuitry 114 such as co-processor(s), clock circuits, cache, power supplies and other well-known circuits. The operation and interrelationship of these various computer components is well-known in the art and does not require further explanation. The display driver 120 may be a video card, printer driver or other common driver software or hardware as required by the output device(s) 106.

The RAM 116 stores the software implementation of the present invention. Typically, the routines of the invention are stored in a mass storage device (not shown) and recalled for temporary storage in the RAM 116 when executed by the CPU 112. In FIG. 1, the invention is embodied in a three-dimensional scene analysis routine 122.

#### A. Parallax-Based Constraints on Pairs of Points

A constraint on the parallax motion of pairs of points between two image frames that represent a three-dimensional scene as imaged by a video camera is described below. The derived constraint can be used to recover a relative 3D



structure invariant of two points from their parallax vectors alone, without any additional information, and in particular, without requiring the recovery of the camera epipoles.

The parallax constraint is extended to multiple frames to form a rigidity constraint on any pair of image points (similar to the trilinear constraint). Namely, inconsistencies in the 3D motion of two points that belong to independently moving 3D objects can be detected based on their parallax displacements among three (or more) frames, without the need to estimate any 3D information.

To derive the parallax constraint, first, the decomposition of the image motion into a homography (i.e., the image motion of an arbitrary planar surface) and residual parallax displacements is described. This decomposition is well-known in the art.

### 1. The Planar Parallax Notations

FIG. 2 provides a geometric interpretation of the planar parallax. Let  $\bar{P} = (X, Y, Z)^T$  and  $\bar{P}^1 = (X^1, Y^1, Z^1)^T$  denote the Cartesian coordinates of a scene point with respect to different camera views, respectively. Let  $\bar{x} = (x, y)^T$  and  $\bar{x}^1 = (x^1, y^1)^T$  denote the corresponding coordinates of the scene point as projected onto the image planes at the two camera positions, respectively. Let  $\bar{p} = (x, y, 1)^T = \frac{1}{Z} P$  and  $\bar{p}^1 = (x^1, y^1, 1)^T = \frac{1}{Z^1} \bar{P}^1$  denote the same point in homogeneous coordinates, respectively.

Let  $S$  be an arbitrary planar surface and  $A'$  the homography that aligns the planar surface  $S$  between the second and first frame (i.e., for all points  $\bar{P} \in S, \bar{P} = A' \bar{P}^1$ ). It can be shown that the image motion can be written (in homogeneous coordinates) as:

$$\bar{p} = \bar{p}_w + k(\bar{p}_w - \bar{e}) \quad (1)$$

In FIG. 2,  $\bar{p}_w$  is the image point in the first frame which corresponds to warping  $\bar{p}$  by the homography  $A' : \bar{p}_w = \frac{A' \bar{p}^1}{a_3^T \bar{p}^1}$  ( $a_3$  being the third row of  $A'$ ). When  $\bar{P}$  is on the 3D planar surface  $S$ , then  $\bar{p}_w = \bar{p}$ . Otherwise, the remaining displacement between  $\bar{p}_w$  and  $\bar{p}$  is proportional to the 3D projective structure ( $k$ )

-8-

+of  $\bar{P}$  with respect to the planar surface  $S$ . Also,  $k = \frac{d_N}{Z} \frac{T_z}{d'}$ , and  $\bar{e} = \frac{1}{T_z} T$  is the epipole in the first frame.  $\bar{T} = (T_x, T_y, T_z)$  is the 3D translation between the two frames as expressed in the coordinate system of the first frame,  $d_N$  is the distance of  $\bar{P}$  from the plane  $S$ , and  $d'$  is the distance to the plane from the second camera center.

*Shape Invariance:* Note that the projective structure  $k$  depends both on the 3D structure and on the camera translation  $\bar{T}$ . However  $\frac{k_1}{k_2} = \frac{d_{N1}}{d_{N2}} \frac{Z_2}{Z_1}$  for any two points  $\bar{P}_1$  and  $\bar{P}_2$  is view-independent, and is therefore a relative structure invariant with respect to camera motion.

When  $T_z = 0$ , Eq. (1) can be rewritten in terms of  $\bar{P}_w$  as:

$$\bar{p} = \bar{p}_w - \tau \bar{T}, \quad (2)$$

where  $\tau = \frac{d_N}{Zd'}$ . Once again,  $\frac{\tau_1}{\tau_2} = \frac{d_{N1}}{d_{N2}} \frac{Z_2}{Z_1}$  provides the same relative structure invariant.

Rewriting Eq. (1) in the form of image displacements yields, in homogeneous coordinates:

$$\bar{p}' - \bar{p} = (\bar{p}' - \bar{p}_w) - k(\bar{p}_w - \bar{e}), \quad (3)$$

and in 2D image coordinates:

$$\bar{u} = (\bar{x}' - \bar{x}) = (\bar{x}' - \bar{x}_w) - k(\bar{x}_w - \bar{x}_e) = \bar{u}_* + \bar{\mu}, \quad (4)$$

where  $\bar{x} = (x, y)^T$  signifies the image point where  $\bar{P} = (x, y, 1)^T$  intersects the image plane of the first frame,  $\bar{x}_e$  signifies the 2D image coordinates of the epipole in the first frame,  $\bar{\mu}_*$  denotes the planar part of the 2D image motion (the homography due to  $S$ ), and  $\bar{\mu}$  denotes the residual parallax 2D motion.

Similarly, when  $T_z = 0$  (Eq. (2)):

$$\bar{u} = (\bar{x}' - \bar{x}) = (\bar{x}' - \bar{x}_w) + t\bar{T} = \bar{u}_* + \bar{\mu}, \quad (5)$$

where  $t = (T_x, T_y)^T$ .

Eqs. (4) and (5) provide the form of parallax notation that will be used in the remaining of the paper. Note that they are expressed in terms of 2D image coordinates. Although we derived the parallax notation differently for  $T_z = 0$  and  $T_z \neq 0$ , they will be unified and treated as a single case in the following sections.

## 2. The Parallax-Based Structure Constraint

Let  $\bar{x}_1 = (x_1, y_1)$  and  $\bar{x}_2 = (x_2, y_2)$  be two image points. We start by analyzing the case when  $T_2 \neq 0$ . From Eq. (4) we have:

$$\bar{\mu}_1 = x_{\bar{w}_1} - \bar{x}_1 = \kappa_1(\bar{x}_* - x_{\bar{w}_1})$$

$$\bar{\mu}_2 = x_{\bar{w}_2} - \bar{x}_2 = \kappa_2(\bar{x}_* - x_{\bar{w}_2}) \quad (6)$$

Let  $\gamma = \frac{d_N}{Z}$  denote the part of the projective structure  $k$  which is point-dependent, i.e.,  $\kappa_i = \gamma_i \frac{T_i}{d}$  ( $i = 1, 2$ ), then:

$$\bar{\mu}_1 = \frac{T_1}{d} \gamma_1 (\bar{x}_* - x_{\bar{w}_1}) \quad (7)$$

$$\bar{\mu}_2 = \frac{T_2}{d} \gamma_2 (\bar{x}_* - x_{\bar{w}_2}) \quad (8)$$

Therefore,

$$\bar{\mu}_1 \gamma_2 - \bar{\mu}_2 \gamma_1 = \gamma_1 \gamma_2 \frac{T_2}{d} (x_{\bar{w}_2} - x_{\bar{w}_1}) \quad (9)$$

This last step eliminated the epipole  $\bar{x}_*$ . Eq. (9) entails that the vectors on both sides of the equation are parallel. Since  $\gamma_1 \gamma_2 \frac{T_2}{d}$  is a scalar, we get:

$$(\bar{\mu}_1 \gamma_2 - \bar{\mu}_2 \gamma_1) \parallel \Delta x_{\bar{w}_2,1},$$

where  $\Delta x_{\bar{w}_2,1} = (x_{\bar{w}_2} - x_{\bar{w}_1})$ . This leads to the pairwise parallax constraint

$$(\bar{\mu}_1 \gamma_2 - \bar{\mu}_2 \gamma_1)^T (\Delta x_{\bar{w}_2,1})_{\perp} = 0,$$

where  $\bar{v}_{\perp}$  signifies a vector perpendicular to  $\bar{v}$ . From Eq. (6),  $x_{\bar{w}_i} = (\bar{x}_i + \bar{\mu}_i)$ ,  $i = 1, 2$ .

Hence, we can rewrite the parallax constraint in terms of the pixel coordinates of the two points and the parallax vectors as:

$$(\bar{\mu}_1 \gamma_2 - \bar{\mu}_2 \gamma_1)^T (\Delta x_{2,1} + \Delta \mu_{2,1})_{\perp} = 0, \quad (10)$$

where  $\Delta x_{2,1} = \bar{x}_2 - \bar{x}_1$ , and  $\Delta \mu_{2,1} = \bar{\mu}_2 - \bar{\mu}_1$ .

Eq. (10) provides a constraint on the structure parameters ( $g_1$  and  $g_2$ ) of the two points, using only their image (pixel) coordinates and their parallax vectors.

$g_1$  and  $g_2$  depend on the second-view. However, if  $\bar{x}_1$  is chosen so that its parallax vector  $\bar{\mu}_1 \neq 0$  (i.e.,  $\bar{P}_1 \notin S$ ), then the parallax constraint can be rewritten in terms of the relative structure invariant (with respect to the first frame) of the two points,  $\frac{\gamma_2}{\gamma_1} = \frac{d_{N1}}{d_{N2}} \frac{Z_2}{Z_1}$ .

$$(\bar{\mu}_1 \frac{\gamma_2}{\gamma_1} - \bar{\mu}_2)^T (\Delta x_{z,1} + \Delta \mu_{z,1})_1 = 0 \quad (11)$$

When  $T_z = 0$ , a constraint stronger than Eq. (11) can be derived:

$$(\bar{\mu}_1 \frac{\gamma_2}{\gamma_1} - \bar{\mu}_2) = 0,$$

however, Eq. (11), still holds. This is important, as we do not have a prior knowledge of  $T_z$  to distinguish between the two cases.

Therefore, given only the parallax vectors of the two image points, their relative structure  $\frac{\gamma_2}{\gamma_1}$  can be recovered from Eq. (11):

$$\frac{\gamma_2}{\gamma_1} = \frac{\mu_2 (\Delta x_{z,1} + \Delta \mu_{z,1})_1}{\bar{\mu}_1^T (\Delta x_{z,1} + \Delta \mu_{z,1})_1},$$

or more concisely,

$$\frac{\gamma_2}{\gamma_1} = \frac{\mu_2 (\Delta x_{w,1})_1}{\bar{\mu}_1^T (\Delta x_{w,1})_1} \quad (12)$$

FIG. 3 displays the constraint geometrically. As shown in this figure,  $\frac{\gamma_2}{\gamma_1} = \frac{AC}{AB}$  where AC and AB are defined in FIG. 3.

The benefit of the constraint in Eq. (12) is that it provides this information directly from the positions and parallax vectors of the two points, without the need to go through the computation of the epipole, using as much information as one point can give on another.

FIG. 4 graphically shows an example of a configuration in which estimating the epipole is very unreliable, whereas, estimating the relative structure directly from Eq. (12) is reliable.

20

### 3. The Parallax-Based Rigidity Constraint

In this section, how the parallax-based structure constraint can be extended to multiple frames, to form a rigidity constraint on pairs of image points that contains neither structure parameters nor camera geometry is described.

Rigidity Over Multiple Frames:

Let  $\bar{x}_1 = (x_1, y_1)$  and  $\bar{x}_2 = (x_2, y_2)$  be two image points in the first (reference) frame. Let  $\bar{\mu}_i^j$  denote the 2D planar parallax motion of  $\bar{x}_i$  from the first frame to frame  $j$ . From Eq. (6) we see that  $x_{\bar{x}_i}^j$  can be computed by  $\bar{x}_i + \bar{\mu}_i^j, (i = 1, 2)$ .

- 5 Using the structure invariance constraint of Eq. (12), for any two frames  $j$  and  $k$  we get:

$$\frac{\gamma_2}{\gamma_1} = \frac{\mu_2 (\Delta x_{\bar{x}_{2,1}})_1}{\bar{\mu}_1^{jT} (\Delta x_{\bar{x}_{2,1}})_1^j} = \frac{\mu_2 (\Delta x_{\bar{x}_{2,1}})_1}{\bar{\mu}_1^{kT} (\Delta x_{\bar{x}_{2,1}})_1^k}.$$

Multiplying by the denominators yields the rigidity constraint of the two points over three frames (reference frame, frame  $j$ , and frame  $k$ ):

$$10 \quad (\bar{\mu}_1^{kT} (\Delta x_{\bar{x}_{1,1}})_1^k) (\bar{\mu}_2^{jT} (\Delta x_{\bar{x}_{2,1}})_1^j) - (\bar{\mu}_1^{jT} (\Delta x_{\bar{x}_{1,1}})_1^j) (\bar{\mu}_2^{kT} (\Delta x_{\bar{x}_{2,1}})_1^k) = 0. \quad (13)$$

- Based on the planar parallax motion trajectory of a single image point (e.g.,  $\bar{x}_1$ ) over several frames, the rigidity constraint (13) states a constraint on the planar parallax motion trajectory of any other point (e.g.,  $\bar{x}_2$ ). The rigidity constraint of Eq. (13) can therefore be applied to detect inconsistencies in the 3D motion of two image points (i.e., indicate whether the two image points are projections of 3D points belonging to a same or different 3D moving object) based on their parallax motion among three (or more) frames alone, without the need to estimate either camera geometry or structure parameters.

Rigidity Over Multiple Points:

- 20 Instead of considering pairs of points over multiple frames, an alternative is to consider multiple points over two frames to come up with a different form of the rigidity constraint.

- Let  $\bar{x}_1 = (x_1, y_1)$ ,  $\bar{x}_2 = (x_2, y_2)$ , and  $\bar{x}_3 = (x_3, y_3)$  be three image points in the first (reference) frame. Let  $\bar{\mu}_i$  denote the 2D planar parallax motion of  $\bar{x}_i$  from the first frame to any other frame, and  $x_{\bar{x}_i} = \bar{x}_i + \bar{\mu}_i, (i = 1, 2, 3)$ .

Using the shape invariance constraint of Eq. (12):

$$\begin{aligned} \frac{\gamma_2}{\gamma_1} &= \frac{\bar{\mu}_2^T (\Delta x_{\bar{x}_{2,1}})_1}{\bar{\mu}_1^T (\Delta x_{\bar{x}_{2,1}})_1}, \\ \frac{\gamma_3}{\gamma_2} &= \frac{\bar{\mu}_3^T (\Delta x_{\bar{x}_{3,2}})_1}{\bar{\mu}_2^T (\Delta x_{\bar{x}_{3,2}})_1}, \end{aligned}$$

-12-

$$\frac{\gamma_3}{\gamma_1} = \frac{\mu_3 (\Delta x_{\tilde{w}_{3,1}})_1}{\bar{\mu}_1^T (\Delta x_{\tilde{w}_{3,1}})_1},$$

Using the equality

$$\frac{\gamma_3}{\gamma_1} = \frac{\gamma_3}{\gamma_2} \frac{\gamma_2}{\gamma_1}$$

and multiplying by the denominators, we get the rigidity constraint for three points over a pair of frames:

$$\bar{\mu}_3^T (\Delta x_{\tilde{w}_{3,1}})_1 \bar{\mu}_2^T (\Delta x_{\tilde{w}_{2,1}})_1 \bar{\mu}_1^T (\Delta x_{\tilde{w}_{1,1}})_1 - \bar{\mu}_2^T (\Delta x_{\tilde{w}_{1,1}})_1 \bar{\mu}_4^T (\Delta x_{\tilde{w}_{2,1}})_1 \bar{\mu}_3^T (\Delta x_{\tilde{w}_{3,1}})_1 = 0 \quad (14)$$

The benefit of the rigidity constraint (14) is in the fact that it provides this information directly from the positions and parallax vectors of the three points, without the need to go through the unstable computation of the epipole, using as much information as two points can give on the third.

### B. Parallax Geometry and an Epipole Dual

In this section, a different way of deriving the parallax-based rigidity constraint is described. Rather than deriving the constraint algebraically, the alternative derivation uses geometry. This leads to a simple and intuitive geometric interpretation of the multiframe rigidity constraint, and to the derivation of a dual point to the epipole. Although this distinct image point (the epipole dual) is pointed out, the rigidity constraint itself does not require the estimation of the dual epipole, just as it did not require the estimation of the epipole itself.

FIG. 6 illustrates the 3D geometric structure associated with the planar parallax of pairs of points between two frames. In this figure,  $S$  is the planar surface, and  $P$  and  $Q$  are the two scene points. As in the case of FIG. 2,  $P_w$  and  $Q_w$  are the intersections of rays  $OP$  and  $OQ$  with the plane  $S$ . Similarly, the points  $\bar{P}_w$  and  $\bar{Q}_w$  on the reference image are the projections of  $P_w$  and  $Q_w$ , and are therefore the points to which the planar homography transforms  $\bar{P}'$  and  $\bar{Q}'$  respectively. In other words,  $\bar{P}_w = A' \bar{P}'$  and  $\bar{Q}_w = A' \bar{Q}'$ . Below,  $\bar{P}_w$  and  $\bar{Q}_w$  are referred to as "warped points."

Let  $R$  be the intersection of the line connecting  $P$  and  $Q$  with the plane  $S$ . Note that the points  $P$ ,  $Q$ ,  $R$ ,  $P_w$  and  $Q_w$  are co-planar. Hence,  $P_w$  and  $Q_w$  and  $R$  are colinear. Of course,  $P$ ,  $Q$  and  $R$  are colinear by constructions.

Let  $\bar{r}$  be the projection of R on the reference image plane. Since  $R \in S, \bar{r} = \bar{r}_w$ . Since  $\bar{p}, \bar{q}, \bar{r}$  and  $\bar{p}_w, \bar{q}_w, \bar{r}$  are the image projection of the colinear warped points P, Q, R, we can infer the following:

$\bar{p}_w, \bar{q}_w$  and  $\bar{r}$  are colinear and  $\bar{p}, \bar{q}$  and  $\bar{r}$  are colinear.

- 5 In other words, the line connecting  $\bar{p}_w$  and  $\bar{q}_w$  and the line connecting  $\bar{p}$  and  $\bar{q}$  intersect at  $\bar{r}$ , the image of the point R.

Note that the point  $\bar{r}$  does not depend on the second camera view. Therefore, if multiple views are considered, then the lines connecting the warped points  $\bar{p}_w^j = A^j \bar{p}^j$  and  $\bar{q}_w^j = A^j \bar{q}^j$  (for any frame j), meet at  $\bar{r}$  for all such views.

- 10 FIG. 7 illustrates the convergence of the lines. Referring to that figure, since the lines qC, pB and rA are parallel to each other and intersect the lines qpr and CAB:

$$\frac{qr}{pr} = \frac{CA}{BA} = \frac{\gamma_q}{\gamma_p}.$$

Similarly,

15 
$$\frac{qr}{pr} = \frac{FD}{ED} = \frac{\gamma_q}{\gamma_p}.$$

Hence

$$\frac{qr}{pr} = \frac{CA}{BA} = \frac{FD}{ED}.$$

- This is the same as the rigidity constraint derived in Eq. (13). Note, however, the rigidity constraint itself does not require the estimation of the point of convergence  $\bar{r}$ , just as it does not require the estimation of the epipole.
- 20

- The point  $\bar{r}$  is the dual of the epipole: the epipole is the point of intersection of multiple parallax vectors between a pair of frames, i.e., the point of intersection of all lines connecting each image point with its warped point between a pair of frames. Whereas, the dual point  $\bar{r}$  is the point of intersection of all lines connecting a pair of points in the reference image and the corresponding pair of warped points from all other frames.
- 25

### C. Applications Of Pairwise Parallax Geometry

In this section, how pairwise parallax geometry, in its various forms, provides an approach to handling some well-known problems in 3D scene

-14-

analysis, in particular: (i) moving object detection, (ii) shape recovery, (iii) new view generation. It is shown that the parallax constraints provide the capability for solving these problems without having to first solve a "more complex" problem.

5        FIG. 5 depicts a flow diagram of a process for performing  
three-dimensional scene analysis which uses the shape constraint of Eq. (12) (or  
equivalently Eq. (11)) and the rigidity constraints of Eqs. (13) and (14). The  
process is represented as an executable software routine 500 that begins at  
step 502 and proceeds to step 504. At step 504, the routine is provided with a  
10    plurality of input images. At step 506, the routine computes a planar parallax  
motion (e.g.,  $\vec{u}$ ) for each point in the image. Then, at step 508, for each motion  
trajectory determined in step 506, one or more of the constraints of Eqs. (11),  
(12), (13), and (14) are applied with respect to all other points. The routine 500  
uses the information from step 508 (e.g., information representing some image  
15    points as being consistent with the constraint and some image points being  
inconsistent with the constraint) within one or more image processing tasks.  
These tasks include, but are not limited to, moving object detection (step 510),  
shape recovery (step 512, and new view generation (step 514). Each of these  
illustrative applications of the inventive technique are described below.

20

### 1. Estimating Planar Parallax Motion

The estimation of the planar parallax motion used for performing the  
experiments presented in this section was done using two successive  
computational steps: (i) 2D image alignment to compensate for a detected  
25    planar motion (i.e., the homography in the form of a 2D parametric  
transformation, and (ii) estimation of residual image displacements between the  
aligned images (i.e., the parallax). Such a system is disclosed in U.S. provisional  
patent application serial number 60/011,496 filed 2/12/96 (Attorney Docket No.  
12040) and herein incorporated by reference.

30

### 2. Shape Recovery

The parallax-based structure constraint (Eq. (12)) can be used to recover a  
3D relative structure between pairs of points directly from their parallax  
vectors. This implies that the structure of the entire scene can be recovered



relative to a single reference image point (with non-zero parallax). Singularities occur when the denominator of the constraint (Eq. (12)) tends to zero, i.e., for points that lie on the line passing through the reference point in the direction of its parallax vector.

5        FIGS. 8a-g show an example of recovering structure of an entire scene relative to a single reference point. Three views obtained by a hand-held camera of a rug covered with toy cars and boxes, were used as the source data whose heights were measured. The detected 2D planar motion was that of the rug (FIG. 8d). A single point with non-zero planar parallax was selected as a  
10        reference point for estimating relative shape (FIG. 8e). FIG. 8f shows the recovered relative structure of the entire scene from two frames (FIGS. 8b and 8c). Regions close to the image boundary were ignored. The obtained results were quite accurate except along the singular line in the direction of the parallax of the reference point. The singular line is evident in FIG. 8f.

15        The singularities can be removed and the quality of the computed structure can be improved either by using multiple frames or by using multiple reference points:

      •Multiple frames: Singularities are removed by using multiple frames if their epipoles are non-colinear. Non-colinearity of epipoles can be detected  
20        through change in the parallax direction of the reference image point.

      •Multiple points: Singularities can be removed by using additional reference image points. An additional reference point should be chosen so that:  
(i) it does not lie on the singular line (i.e., in the direction of the parallax vector) of the first reference point (it should preferably be chosen on the line  
25        perpendicular to that), and (ii) the additional reference point should be first verified to move consistently with the first reference point through the rigidity constraint of Eq. (13) over a few frames.

      Combinations of multiple reference points over multiple frames can also be used. FIG. 8g shows an example of recovering structure of an entire scene  
30        from three frames relative to the same single reference point as in FIG. 8f. The singular line in FIG. 8f has disappeared.

      The ability to obtain relatively good structure information even with respect to a single point has several important virtues:

- It does not require the estimation of the epipole, and therefore, does not require dense parallax information.

- Unlike conventional techniques for recovering structure, it provides the capability to handle dynamic scenes, as it does not require having a collection of image points which is known a priori to belong to the single 3D moving object.

- Since it relies on a single parallax vector, it provides a natural continuous way to bridge the gap between 2D cases, that assume only planar motion exists, and 3D cases that rely on having parallax data.

10 3. Moving Object Detection

A number of techniques exist to handle multiple motions analysis in the simpler 2D case, where motions of independent moving objects are modeled by 2D parametric transformation. These methods, however, detect points with planar parallax motion as moving objects, as they have a different 2D image motion than the planar part of the background scene.

In the general 3D case, the moving object detection problem is much more complex, since it requires detecting 3D motion inconsistencies. Typically, this is done by recovering the epipolar geometry. Trying to estimate epipolar geometry (i.e., camera motion) in the presence of multiple moving objects, with no prior segmentation, is extremely difficult. This problem becomes even more acute when there exists only sparse parallax information.

FIG. 9a graphically displays an example of a configuration in which estimating the epipole in the presence of multiple moving objects can produce relatively large errors, even when using clustering techniques in the epipole domain as suggested by some conventional techniques. Relying on the epipole computation to detect inconsistencies in 3D motion fails in detecting moving objects in such cases.

In FIG. 9a, the camera is translating to the right. The only static object with pure parallax motion is that of the tree. The ball is falling independently. The epipole may be incorrectly computed as  $e$ . The false epipole  $e$  is consistent with both motions.

The parallax rigidity constraint (Eq. 13) can be applied to detect inconsistencies in the 3D motion of one image point relative to another directly from their "parallax" vectors over multiple (three or more) frames, without the

need to estimate either camera geometry or shape parameters. This provides a useful mechanism for clustering (or segmenting) the "parallax" vectors (i.e., the residual motion after planar registration) into consistent groups belonging to consistently 3D moving objects, even in cases such as in FIG. 9a where the parallax information is minimal, and the independent motion is not negligible. FIG. 9b graphically illustrates how the rigidity constraint of Eq. (13), when applied, detects the 3D inconsistency over three frames.

FIGS. 10a-f shows an example of using the rigidity constraint of Eq. (13) to detect 3D inconsistencies. In this sequence, the camera is in motion (translating from left to right), inducing parallax motion of different magnitudes on the house, road, and road-sign. The car moves independently from left to right. The detected 2D planar motion was that of the house. The planar parallax motion was computed after 2D registration of the three images with respect to the house (FIG. 10d). A single point on the road-sign was selected as a point of reference (FIG. 10e). FIG. 10f displays the measure of inconsistency of each point in the image with respect to the selected road-sign point. Bright regions indicate large values when applying the constraint of Eq. (13) (i.e., violations in 3D rigidity detected over three frames with respect to the road-sign point). The region which was detected as moving 3D-inconsistently with respect to the road-sign point corresponds to the car. Regions close to the image boundary were ignored. All other regions of the image were detected as moving 3D-consistently with the road-sign point. Therefore, assuming an uncalibrated camera, this method provides a mechanism for segmenting all non-zero residual motion vectors (after 2D planar stabilization) into groups moving consistently (in the 3D sense).

FIG. 11a-f shows another example of using the rigidity constraint of Eq. (13) to detect 3D inconsistencies. In this sequence, the camera is mounted on a helicopter flying from left to right, inducing some parallax motion (of different magnitudes) on the house-roof and trees (bottom of the image), and on the electricity poles (by the road). Three cars move independently on the road. The detected 2D planar motion was that of the ground surface (FIG. 11d). A single point was selected on a tree as a point of reference (FIG. 11e). FIG. 11f displays the measure of inconsistency of each point in the image with respect to the selected reference point. Bright regions indicate 3D-inconsistency detected over

-18-

three frames. The three cars were detected as moving inconsistently with the selected tree point. Regions close to the image boundary were ignored. All other image regions were detected as moving consistently with the selected tree point.

5 In the prior art, a rigidity constraint between three frames in the form of a trilinear tensor has been presented using regular image displacements. However, it requires having a collection of a set of image points which is known *a priori* to belong to the single 3D moving object. Selecting an inconsistent set of points leads to an erroneous tensor, and, hence, false moving object detection.

10 The ability of the parallax rigidity constraint of the present invention to detect 3D-inconsistency with respect to a single point, provides a natural way to bridge between 2D algorithms (which assume that any 2D motion different than the planar motion is an independently moving object), and 3D algorithms (which rely on having prior knowledge of a consistent set of points, or alternatively, dense parallax data).

15

#### 4. New View Generation

This section describes an approach based on the parallax rigidity constraint for generating novel views using a set of "model" views.

20 Methods for generating new views based on recovering epipolar geometry are likely to be more noise sensitive than methods that generate the new view based on 2D information alone, i.e., without going from 2D through a 3D medium in order to reproject the information once again onto a new 2D image plane (the virtual view). The approach described below for new view generation does not require any epipolar geometry or shape estimation.

25 Given two "model" frames, planar parallax motion can be computed for all image points between the first (reference) frame and the second frame. An image point with non-zero parallax is selected, and a "virtual" parallax vector is defined for that point from the reference frame to the "virtual" frame to be generated. The rigidity constraint (Eq. 13) then specifies a single constraint on  
30 the virtual parallax motion of all other points from the reference frame to the virtual frame. Since each 2D parallax vector has two components (i.e., two unknowns), at least two "virtual" parallax vectors are needed to be specified in order to solve for all other virtual parallax vectors. Once the virtual parallax vectors are computed, the new virtual view can be created by warping the

reference image twice: First, warping each image point by its computed virtual parallax. Then, globally warping the entire frame with a 2D virtual planar motion for the virtual homography.

Note that two virtual parallax vectors may not provide sufficient constraints for some image points. This is due to unfavorable location of those points in the image plane with respect to the two selected reference points and their parallax vectors. However, other image points, for whom the constraint was robust and sufficient to produce reliable virtual parallax, can be used (once their virtual parallax has been computed) as additional points to reliably constrain the virtual parallax of the singular points.

#### 10 D. The Generalized Parallax Constraint

In this section, how the pairwise-parallax constraint (Eqs. (11, 12, 13, and 14)) can be extended to handle full image motion (as opposed to parallax motion), even when the homography is unknown, is described. This is useful for handling scenes that do not contain a physical planar surface. A form of a generalized parallax constraint between two frames in terms of the unknown homography parameters and the relative projective structure of pairs of points is described.

Eqs. (1) and (2) can be unified into a single form as:

$$\bar{p} = \bar{p}_w + \tau(T_z \bar{p}_w - \bar{T}) = \frac{A' \bar{p}'}{a'_{3p}} + \tau(T_z \frac{A' \bar{p}'}{a'_{3p}} - \bar{T}). \quad (15)$$

where  $\tau = \frac{d_N}{Zd'} = \gamma \frac{1}{d}$  (see notations above).  $A'$  is an unknown homography from frame 2 to the first frame. It could relate to any planar surface in the scene, in particular a virtual plane. Importantly, before factoring out the structure, we first factor out the epipole, in a similar manner performed above.

Let  $\bar{P}_1$  and  $\bar{P}_2$  be two scene points with homogeneous coordinates  $\bar{p}_1$  and  $\bar{p}_2$ . Then:

$$\frac{1}{\tau_1} (\bar{p}_1 - \frac{A' \bar{p}_1}{a'_{3p_1}}) = T_z \frac{A' \bar{p}_1}{a'_{3p_1}} - \bar{T}$$

$$25 \quad \frac{1}{\tau_2} (\bar{p}_2 - \frac{A' \bar{p}_2}{a'_{3p_2}}) = T_z \frac{A' \bar{p}_2}{a'_{3p_2}} - \bar{T}$$

Subtracting the two equations eliminates  $\bar{T}$ , yielding:

$$\frac{1}{\tau_1} \left( \bar{p}_1 - \frac{A' \bar{p}_1}{a'_3 \bar{p}_1} \right) - \frac{1}{\tau_2} \left( \bar{p}_2 - \frac{A' \bar{p}_2}{a'_3 \bar{p}_2} \right) = T_2 \left( \frac{A' \bar{p}_1}{a'_3 \bar{p}_1} - \frac{A' \bar{p}_2}{a'_3 \bar{p}_2} \right). \quad (16)$$

The equality (16) entails that the vectors on both sides of the equation are parallel, which leads to:

$$\left( \frac{1}{\tau_1} \left( \bar{p}_1 - \frac{A' \bar{p}_1}{a'_3 \bar{p}_1} \right) - \frac{1}{\tau_2} \left( \bar{p}_2 - \frac{A' \bar{p}_2}{a'_3 \bar{p}_2} \right) \right)^T \left( \frac{A' \bar{p}_1}{a'_3 \bar{p}_1} - \frac{A' \bar{p}_2}{a'_3 \bar{p}_2} \right)_\perp = 0.$$

Multiplying both sides by  $\tau_2$  gives the generalized parallax constraint in terms the relative projective structure  $\frac{\tau_2}{\tau_1} = \frac{\gamma_2}{\gamma_1}$ :

$$\left( \frac{\gamma_2}{\gamma_1} \left( \bar{p}_1 - \frac{A' \bar{p}_1}{a'_3 \bar{p}_1} \right) - \left( \bar{p}_2 - \frac{A' \bar{p}_2}{a'_3 \bar{p}_2} \right) \right)^T \left( \frac{A' \bar{p}_1}{a'_3 \bar{p}_1} - \frac{A' \bar{p}_2}{a'_3 \bar{p}_2} \right)_\perp = 0. \quad (17)$$

The generalized parallax constraint (17) is expressed in terms of the homography  $A'$ , the image coordinates of a pair of points in two frames, and the relative projective structure of the two points. The generalized constraint does not involve the epipoles.

The generalized parallax constraint suggests a new implicit representation of general 2D image motion: Rather than looking at the representation of 2D image motion in terms of: homography plus epipole plus projective structure, it suggests an implicit representation of 2D image motion in terms of homography plus relative projective structure of pairs of points. Since this representation does not contain the epipole, it can be easily extended to multiple frames.

Using Eq. (17), the invariant relative projective structure  $\frac{\gamma_2}{\gamma_1}$  can be expressed explicitly in terms of the plane homography and the image positions of the two points in the two frames:

$$\frac{\gamma_2}{\gamma_1} = \frac{\left( \bar{p}_2 - \frac{A' \bar{p}_2}{a'_3 \bar{p}_2} \right)^T \left( \frac{A' \bar{p}_1}{a'_3 \bar{p}_1} - \frac{A' \bar{p}_2}{a'_3 \bar{p}_2} \right)_\perp}{\left( \bar{p}_1 - \frac{A' \bar{p}_1}{a'_3 \bar{p}_1} \right)^T \left( \frac{A' \bar{p}_1}{a'_3 \bar{p}_1} - \frac{A' \bar{p}_2}{a'_3 \bar{p}_2} \right)_\perp} \quad (18)$$

Since the computations are invariant to new camera positions, it can be factored out over multiple frames ( $j$  and  $k$ ):

$$\frac{(\bar{p}_2 - \frac{A^j p_2^j}{a_3^j p_2^j})^T (\frac{A^j p_1^j}{a_3^j p_1^j} - \frac{A^j p_2^j}{a_3^j p_2^j})_\perp}{(\bar{p}_1 - \frac{A^j p_1^j}{a_3^j p_1^j})^T (\frac{A^j p_1^j}{a_3^j p_1^j} - \frac{A^j p_2^j}{a_3^j p_2^j})_\perp} = \frac{(\bar{p}_2 - \frac{A^k p_2^k}{a_3^k p_2^k})^T (\frac{A^k p_1^k}{a_3^k p_1^k} - \frac{A^k p_2^k}{a_3^k p_2^k})_\perp}{(\bar{p}_1 - \frac{A^k p_1^k}{a_3^k p_1^k})^T (\frac{A^k p_1^k}{a_3^k p_1^k} - \frac{A^k p_2^k}{a_3^k p_2^k})_\perp},$$

which leads to a rigidity constraint of the form:

$$\begin{aligned} & (\bar{p}_2 - \frac{A^j p_2^j}{a_3^j p_2^j})^T (\frac{A^j p_1^j}{a_3^j p_1^j} - \frac{A^j p_2^j}{a_3^j p_2^j})_\perp (\bar{p}_1 - \frac{A^k p_1^k}{a_3^k p_1^k})^T (\frac{A^k p_1^k}{a_3^k p_1^k} - \frac{A^k p_2^k}{a_3^k p_2^k})_\perp - \\ & (\bar{p}_2 - \frac{A^k p_2^k}{a_3^k p_2^k})^T (\frac{A^k p_1^k}{a_3^k p_1^k} - \frac{A^k p_2^k}{a_3^k p_2^k})_\perp (\bar{p}_1 - \frac{A^j p_1^j}{a_3^j p_1^j})^T (\frac{A^j p_1^j}{a_3^j p_1^j} - \frac{A^j p_2^j}{a_3^j p_2^j})_\perp = 0. \end{aligned} \quad (19)$$

5

Eq. (19) is a rigidity constraint on a pair of points across three frames. Like the trilinear tensor of the prior art, it involves the parameters of two homographies across three frames. Unlike the trilinear tensor, it does not contain the epipole, but instead, is expressed on pairs of points.

10

The trilinear constraints are based on a initial reference points, and any additional point adds four linearly independent equations to constrain the unknowns of the tensor (which are combinations of the homography parameters and the epipoles).

15

In the generalized parallax rigidity constraint, the basis is a pair of points. Here also, any additional point adds four linearly independent rigidity constraints. These can be derived through factoring out  $TZ$  from Eq. (16) with the additional third point (still within a pair of frames) to form the four linearly independent equations over three frames.

Although various embodiments which incorporate the teachings of the present invention have been shown and described in detail herein, those skilled in the art can readily devise many other varied embodiments that still incorporate these teachings.

20

-22-

## We Claim:

- 1           1. Method for image processing comprising the steps of:
  - 2           a) receiving a plurality of two dimensional images representative of a scene;
  - 3           b) computing a parallax-related constraint for a pair of points within the plurality of
  - 4 images;
  - 5           c) applying the parallax-related constraint to a plurality of points within the plurality
  - 6 of images in order to generate information representative of whether a given point within
  - 7 the plurality of images is consistent with the parallax-related constraint; and
  - 8           d) using the generated information for an image processing task related to the
  - 9 received plurality of images.
- 1           2. Method for image processing to detect a moving object comprising the steps of:
  - 2           a) receiving a plurality of two dimensional images representative of a scene;
  - 3           b) computing a parallax rigidity constraint for at least a pair of points within the
  - 4 plurality of images;
  - 5           c) applying the parallax rigidity constraint to a plurality of points within the
  - 6 plurality of images in order to generate information representative of whether a given point
  - 7 within the plurality of images is consistent with the parallax rigidity constraint; and
  - 8           d) using the generated information to perform moving object detection related to the
  - 9 received plurality of images.
- 1           3. The method of claim 2, wherein the parallax rigidity constraint derives from
- 2 position information of the pair of points and parallax vectors of the pair of points for three
- 3 images.
- 1           4. The method of claim 2, wherein the parallax rigidity constraint derives from
- 2 position information of at least three points and parallax vectors of the at least three points
- 3 for two images.
- 1           5. Method for image processing to recover scene structure comprising the steps of:
  - 2           a) receiving a plurality of two dimensional images representative of a scene;



3           b) computing a parallax structure constraint for a pair of points within the plurality  
4 of images;

5           c) applying the parallax structure constraint to a plurality of points within the  
6 plurality of images in order to generate information representative of whether a given point  
7 within the plurality of images is consistent with the parallax structure constraint; and

8           d) using the generated information to perform recovery of scene structure related to  
9 the received plurality of images.

1           6. Method for image processing to generate new views comprising the steps of:

2           a) receiving a plurality of two dimensional images representative of a scene;

3           b) computing a parallax rigidity constraint for a pair of points within the plurality of  
4 images;

5           c) applying the parallax rigidity constraint to a plurality of points within the  
6 plurality of images in order to generate information representative of whether a given point  
7 within the plurality of images is consistent with the parallax rigidity constraint; and

8           d) using the generated information to perform new view generation related to and  
9 based on the received plurality of images.

1           7. The method of claim 6, wherein the parallax rigidity constraint derives from  
2 position information of the pair of points and parallax vectors of the pair of points for three  
3 images.

1           8. The method of claim 6, wherein the parallax rigidity constraint derives from  
2 position information of at least three points and parallax vectors of the at least three points  
3 for two images.

1           9. Apparatus for image processing comprising:

2           a source of two dimensional images representative of a scene;

3           a computer processor for processing the two dimensional images including:

4                 means for receiving a plurality of two dimensional images representative of  
5 a scene;

6 means for computing a parallax-related constraint for a pair of points within  
7 the plurality of images;

8 means for applying the parallax-related constraint to a plurality of points  
9 within the plurality of images in order to generate information representative of whether a  
10 given point within the plurality of images is consistent with the parallax-related constraint;

11 means for using the generated information for an image processing task  
12 related to the received plurality of images and generating an output related thereto;

13 an output device for presenting the output of the image processing task.

1 10. The apparatus of claim 9, wherein the source of images includes a video  
2 camera.

1 11. A computer-readable medium having stored thereon a plurality of instructions,  
2 the plurality of instructions including instructions which, when executed by a processor,  
3 cause the processor to perform the steps of:

4 a) receiving, in digital form, a plurality of two dimensional images representative of  
5 a scene;

6 b) computing a parallax-related constraint for a pair of points within the plurality of  
7 images;

8 c) applying the parallax-related constraint to a plurality of points within the plurality  
9 of images in order to generate information representative of whether a given point within  
10 the plurality of images is consistent with the parallax-related constraint; and

11 d) using the generated information for an image processing task related to the  
12 received plurality of images.

## AMENDED CLAIMS

[received by the International Bureau on 2 July 1997 (02.07.97);  
original claims 1-11 replaced by new claims 1-17 (5 pages)]

- 1 1. A method for image processing comprising the steps of:
  - 2 a) receiving a plurality of two dimensional images representative of a scene;
  - 3 b) computing a parallax related constraint for a pair of points within the plurality of
  - 4 images, said parallax related constraint being independent of any epipolar geometry which
  - 5 may be defined for the pair of points;
  - 6 c) applying the parallax related constraint to a plurality of points within the plurality of
  - 7 images in order to generate information representative of whether a given point within the
  - 8 plurality of images is consistent with the parallax related constraint; and
  - 9 d) using the generated information for an image processing task related to the received
  - 10 plurality of images.
- 1 2. A method for image processing to detect moving objects comprising the steps of:
  - 2 a) receiving a plurality of two dimensional images representative of a scene;
  - 3 b) computing a parallax rigidity constraint for at least a pair of points within the
  - 4 plurality of images said parallax rigidity constraint being independent of any epipolar
  - 5 geometry which may be defined for the pair of points;
  - 6 c) applying the parallax constraint to a plurality of points within the plurality of images
  - 7 in order to generate information representative of whether a given point within the plurality of
  - 8 images is consistent or inconsistent with the parallax constraint; and
  - 9 d) using the generated information to perform moving object detection related to the
  - 10 received plurality of images.
- 1 3. The method of claim 2, wherein the parallax rigidity constraint is derived from
- 2 position information of the pair of points and parallax vectors of the pair of points generated
- 3 from three of the received plurality of images.
- 1 4. The method of claim 2, wherein the parallax rigidity constraint is derived from
- 2 position information of at least three points and parallax vectors of the at least three points
- 3 generated from two of the received plurality of images.
- 1 5. The method of claim 2 wherein:
  - 2 step (c) further comprises the steps of:
  - 3 (c1) applying a two dimensional transformation to said plurality of images to align
  - 4 regions of said plurality of images and identify a plurality of misaligned regions of said
  - 5 plurality of images;

6 (c2) segmenting and aligning said identified regions to produce segmented regions; and  
7 step (d) further comprises the step of  
8 (d1) iteratively aligning and segmenting the segmented regions until a remaining region  
9 fulfills a criteria that identifies the remaining region as a residual motion within said plurality  
10 of images.

1 6. The method of claim 5 wherein the criteria fulfilled by the remaining region  
2 includes inconsistency with the parallax rigidity constraint.

1 7. The method of claim 2 wherein:  
2 step (b) comprises the steps of:

3 (b1) applying a two dimensional transformation to said plurality of images to  
4 align regions of said plurality of images and identify a plurality of misaligned regions  
5 of said plurality of images; and

6 step (c) comprises the steps of:

7 (c1) segmenting said misaligned regions to produce segmented regions;

8 (c2) iteratively aligning and segmenting the misaligned regions until a  
9 remaining misaligned region fulfills a criteria that identifies the remaining misaligned  
10 region as motion within said plurality of images.

1 8. A method for image processing to recover scene structure comprising the steps of:

2 a) receiving a plurality of two dimensional images representative of a scene;

3 b) computing a parallax structure constraint for a pair of points within the plurality of  
4 images said parallax structure constraint being independent of any epipolar geometry which  
5 may be defined for the pair of points;

6 c) applying the parallax structure constraint to a plurality of points within the plurality  
7 of images in order to generate information representative of whether a given point within the  
8 plurality of images is consistent with the parallax structure constraint; and;

9 d) using the generated information to perform recovery of scene structure related to the  
10 received plurality of images.

1 9. A method for image processing to generate new views comprising the steps of:

2 a) receiving a plurality of two dimensional images representative of a scene;

3 b) computing a planar parallax motion trajectory for a subset of points within the  
4 plurality of images to generate information representative of relative parallax motion among

5 the subset points said planar parallax motion trajectory being independent of any epipolar  
6 geometry which may be defined for the subset of points;

7 c) selecting at least two points from the subset of points having non-zero parallax  
8 motion; and

9 d) defining respective virtual parallax motion vectors for each of the selected points,  
10 the virtual parallax motion vectors representing respective locations of the selected points in a  
11 virtual scene;

12 e) applying a parallax rigidity constraint to at least the subset of points within the  
13 plurality of images with reference to the defined parallax motion vectors to generate a parallax  
14 motion vector for each of the subject of points;

15 f) warping each point in the plurality of images having a parallax motion vector by  
16 that parallax motion vector to generate an intermediate virtual scene; and

17 g) globally warping the intermediate virtual scene with a 2-D virtual planar motion  
18 corresponding to the virtual scene to produce the virtual scene.

1 10. The method of claim 9, wherein the parallax rigidity constraint is derived from  
2 position information of the at least two points and the parallax vectors of the at least two  
3 points generated from three of the received plurality of images.

1 11. The method of claim 9, wherein the parallax rigidity constraint is derived from  
2 position information of at least three points and parallax vectors of the at least three points  
3 generated from two of the received plurality of images.

1 12. Apparatus for image processing comprising:

2 a source of two dimensional images representative of a scene;

3 a computer processor for processing the two dimensional images including:

4 (a) means for receiving a plurality of two dimensional images representative of  
5 a scene;

6 (b) means for applying a parallax-related constraint to a plurality of points  
7 within the plurality of images to generate information representative of whether a given  
8 point within the plurality of images is consistent with the parallax-related constraint,  
9 said parallax-related constraint being independent of any epipolar geometry which may  
10 be defined for the plurality of points;

11 (c) means for applying a parallax constraint to a plurality of points within the  
12 plurality of images in order to generate information representative of whether a given  
13 point within the plurality of images is consistent with the parallax-related constraint;

14 (d) means for using the generated information for an image processing task  
15 related to the received plurality of images and generating an output signal related  
16 thereto;

17 (e) an output device for presenting the output signal of the image processing  
18 task.

1 13. The apparatus of claim 12, wherein the source of images includes a video camera.

1 14. Apparatus for detecting object motion within a sequence of images representing a  
2 scene comprising:

3 means for selecting a plurality of images from said sequence of images;

4 means for applying a parallax constraint to said plurality of images to align regions of  
5 said plurality of images and identify a misaligned region of said plurality of images;

6 means for identifying parallax motion within said misaligned region; and

7 means for removing parallax motion from the misaligned region to detect a moving  
8 object within the scene.

1 15. The apparatus of claim 14 wherein said means for removing parallax motion  
2 further comprises means for identifying image components that are moving consistently with  
3 respect to a parallax motion constraint to identify parallax generated motion within residual  
4 motion in the scene.

1 16. A computer-readable medium having stored thereon a plurality of instructions, the  
2 plurality of instructions including instructions which, when executed by a processor, cause the  
3 processor to perform the steps of:

4 a) receiving, in digital form, a plurality of two dimensional images representative of a  
5 scene;

6 b) computing a parallax-related constraint for a pair of points within the plurality of  
7 images said parallax-related constraint being independent of any epipolar geometry which may  
8 be defined for the pair of points;

9 c) applying the parallax-related constraint to a plurality of points within The plurality  
10 of images in order to generate information representative of whether a given point within the  
11 plurality of images is consistent with the parallax-related constraint; and

12           d) using the generated information for an image processing task related to the received  
13 plurality of images.

1           17. A computer readable medium having stored thereon a plurality of instructions  
2 including instructions which, when executed by a processor, cause the processor to perform  
3 the steps of:

4           (a) selecting a plurality of images from said sequence of images;

5           (b) applying a two dimensional transformation to said plurality of images to align  
6 stationary regions of said plurality of images and identify a misaligned region of said plurality  
7 of images;

8           (c) determining residual motion within said misaligned region; and

9           (d) removing parallax motion from said residual motion to detect a moving object  
10 within the scene.

1

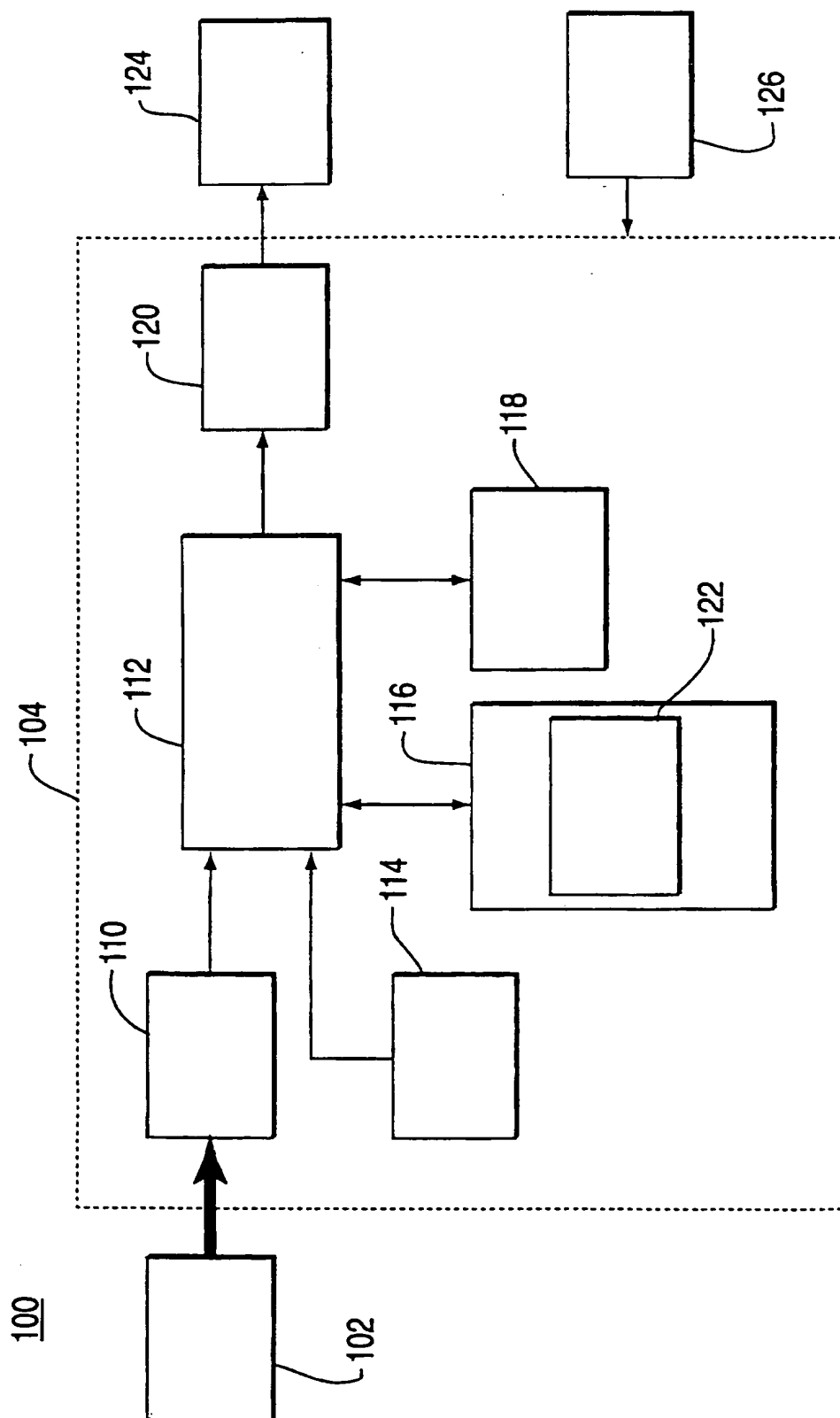
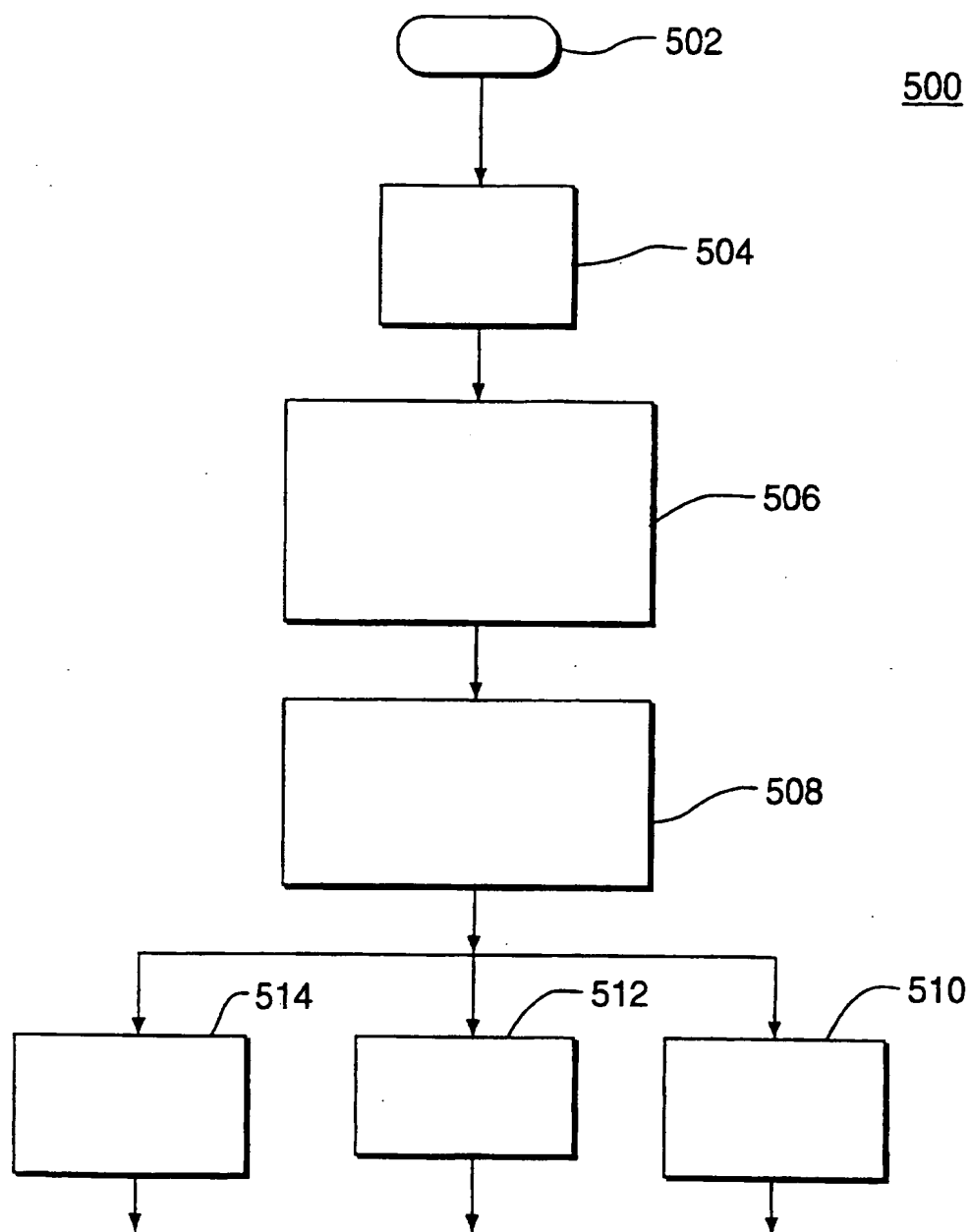
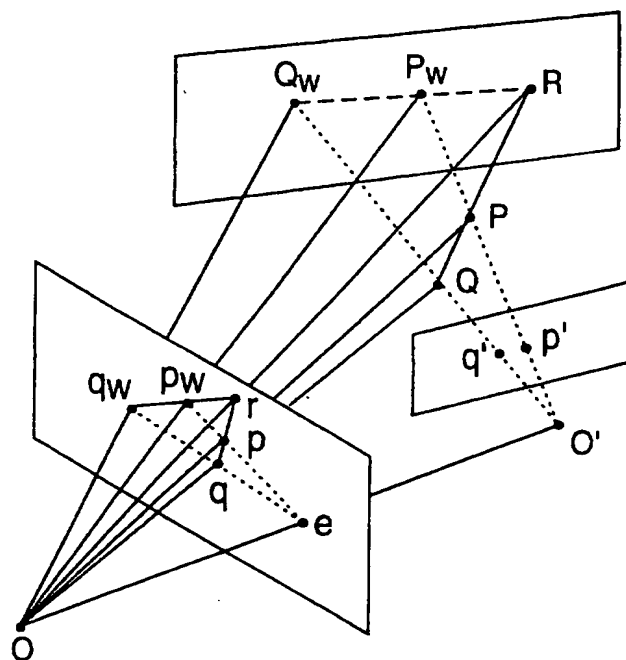


FIG. 1

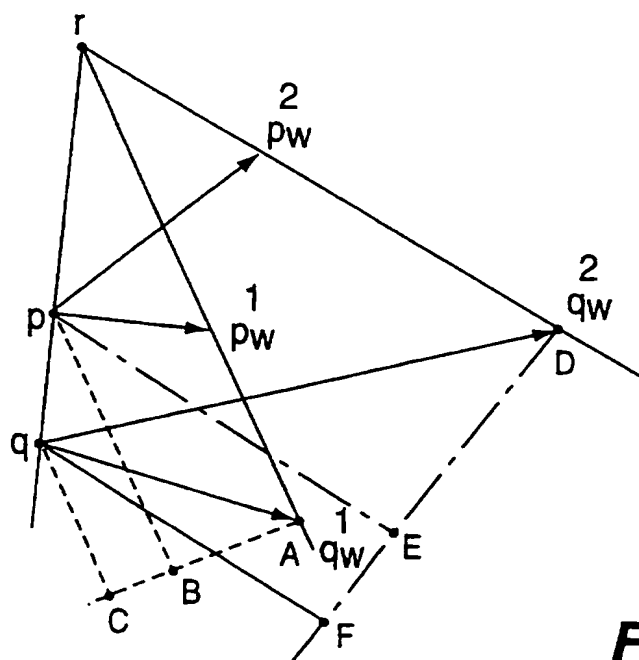




**FIG. 5**



**FIG. 6**



**FIG. 7**

5/8

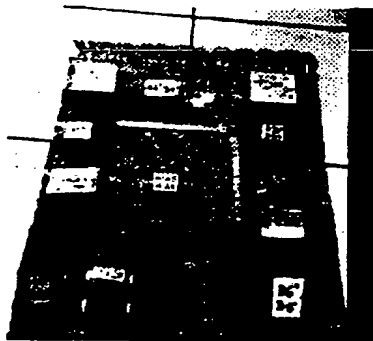


Fig. 8a

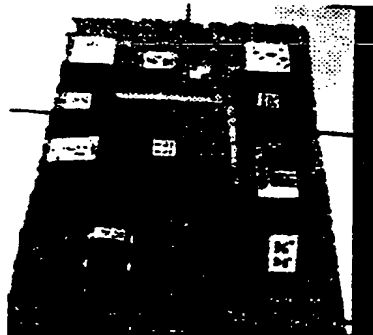


Fig. 8b

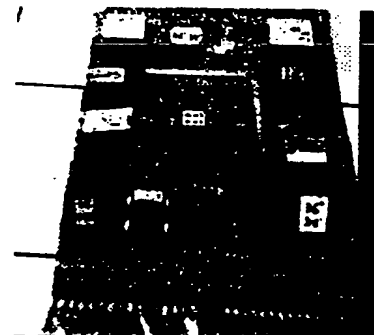


Fig. 8c

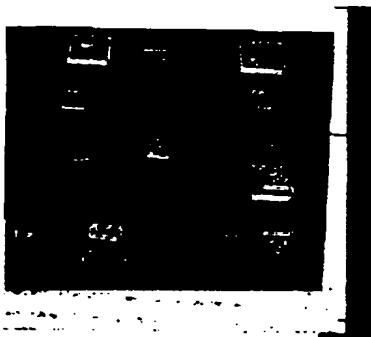


Fig. 8d

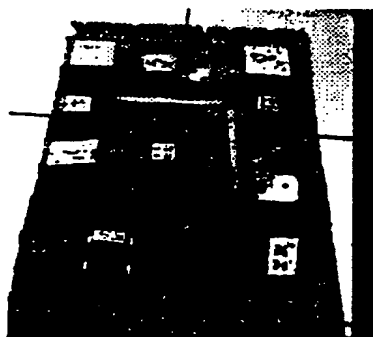


Fig. 8e



Fig. 8f

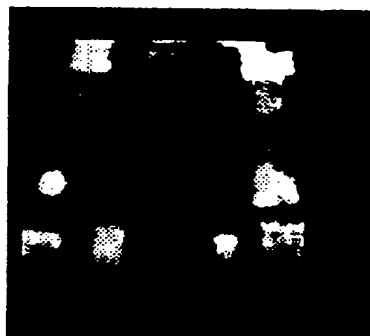
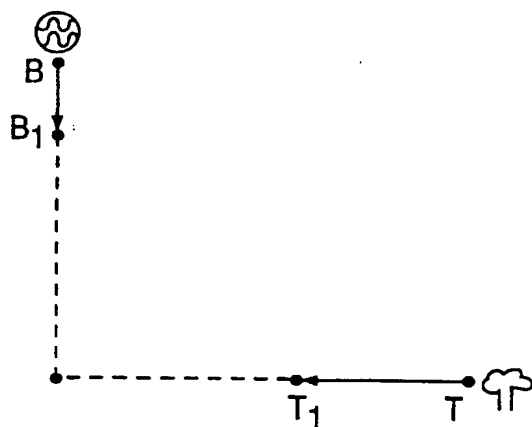
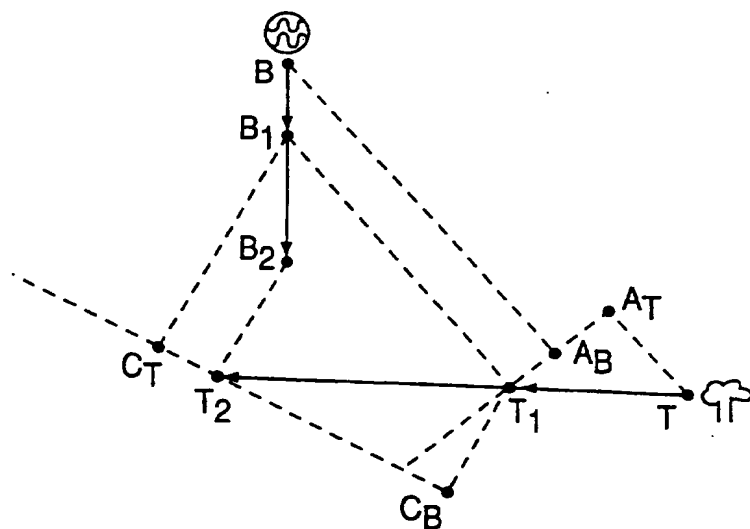


Fig. 8g

**FIG. 9a****FIG. 9b**

7/8



Fig. 10a



Fig. 10b

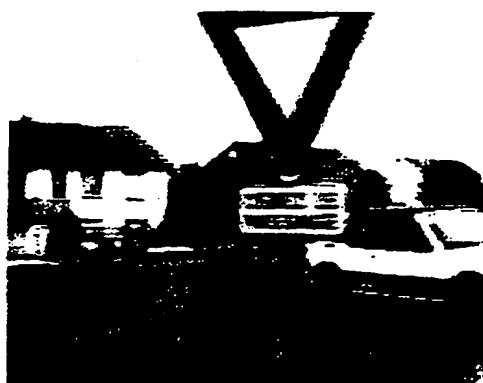


Fig. 10c



Fig. 10d

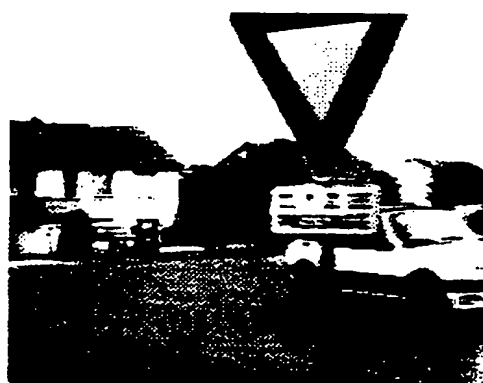


Fig. 10e



Fig. 10f



Fig. 11a

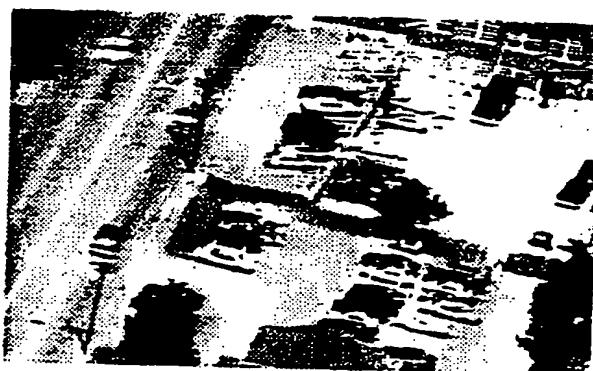


Fig. 11b

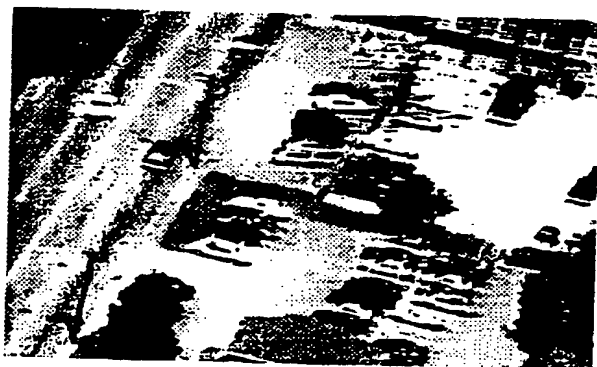


Fig. 11c



Fig. 11d



Fig. 11e

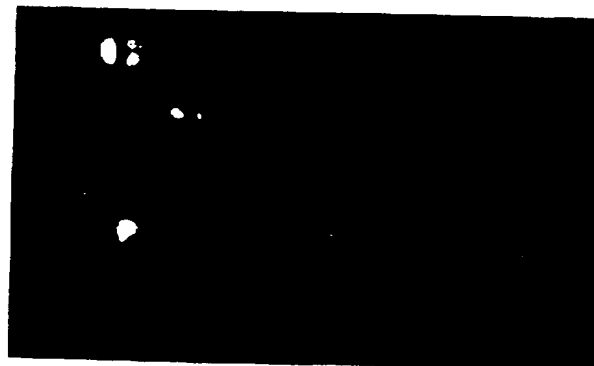


Fig. 11f

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/02115

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G 01 B 11/04

US CL : 382/154

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : Please See Extra Sheet.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

IEEE, COMPENDEX

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	US 4,654,872 A (HISANO ET AL) 31 March 1987, col 2, line 26 - col. 3, line 41	1, 5-11 ----- 2-4
X -- Y	US 4,573,191 A (KIDODE et al) 25 February 1986, col. 2, line 29 - col. 3, line 45	1, 5-11 ----- 2-4
Y	US 5,473,364 A (BURT) 05 December 1995, col. 5, line 53 - col. 6, line 18	2-4
A	US 5,586,231 A (FLORENT et al) 17 December	
A	US 5,577,130 A (WU) 19 November 1996	1-11
A	US 5,249,035 A (YAMANAKA) 28 September 1993	1-11



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"G" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

24 APRIL 1997

Date of mailing of the international search report

06 JUN 1997

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

BRAIN P. WERNER

Telephone No. (703) 308-5876



# INTERNATIONAL SEARCH REPORT

International Application No.  
PCT/US97/02115

## B. FIELDS SEARCHED

Minimum documentation searched

Classification System: U.S.

382/103, 104, 106, 107, 154, 239, 291; 348/20, 25, 26, 154, 155, 169; 356/12, 27, 144, 145, 146, 385; 359/458;  
364/565; 395/125, 136